

**AGRI-OCEAN DSPACE:
FAO AND UNESCO-IOC/IODE COMBINE EFFORTS IN THEIR SUPPORT OF OPEN
ACCESS**

Marc Goovaerts⁽¹⁾, Denys Slipetskyy⁽²⁾, Imma Subirats⁽³⁾ and Sarah Dister⁽³⁾

Addresses: (1) marc.goovaerts@uhasselt.be, Hasselt University, Belgium
(2) d.slipetskyy@ibss.org.ua, Institute of Biology of the Southern Seas, Ukraine
(3) Imma.Subirats@fao.org, FAO of the United Nations

*7th IAMSILIC Conference
Zanzibar, Tanzania, October 16-20, 2011*

Abstract: The United Nations agencies of FAO and UNESCO-IOC/IODE have created in a joint initiative a customized version of DSpace using standards and controlled vocabularies in oceanography, marine science, food, agriculture, development, fisheries, forestry, natural resources and related sciences.¹

Keywords: United Nations, oceanography, marine science, food, agriculture, fisheries, forestry, natural resources.

The collaboration between FAO and UNESCO-IOC/IODE has the following goals:

- Promoting open access to scientific information on the topics of food, agriculture, development, fisheries, forestry, natural resources and related sciences for FAO and oceanography and marine sciences for UNESCO-IOC/IODE.
- Ensuring the metadata quality of repositories and the use of thesauri and other forms of authority control.
- Contributing to the development of sustainable repositories by the use of tools to make scientific publications (and later data) more accessible and visible.
- Removing access barriers by encouraging the creation of new service providers based on existing and mature metadata and semantics technology.

AgriOcean DSpace is based on OceanDocs, a customization of DSpace. It integrates the alternative document- type-based submission module developed in DSpace 1.4.2 . This version already supported a more refined metadata and integrated the ASFA Thesaurus in the submission module. Other features were a batch import module on browser level for RIS and Bibtex files.

These developments have been ported to DSpace, version 1.7. It uses new features, in the first place the authority control, statistics and harvesting modules.

The developments of FAO on AGRIS AP and thesaurus integration do complete AgriOcean DSpace. Finally an easy-to-install Windows-based version of AgriOcean DSpace is available since August 15 2011.

1. Historical background

The OceanDocs Network

OceanDocs started in 2004 as a project in the framework of OdinAfrica. In 2007 the OdinPubAfrica repository became OceanDocs, a repository for the IODE-related communities. Also some of the IODE partners were interested in setting-up their own repository. OceanDocs is now the repository network of UNESCO/IOC-IODE.²

¹ <http://aims.fao.org/tools/agrioccean-dspace>

² M. Goovaerts, OCEANDocs and Open Science Directory: two facets of the information policy of UNESCO/IOC-IODE, IAMSILIC Conference Proceedings 2009, <http://hdl.handle.net/1912/3777>, p. 1-8.

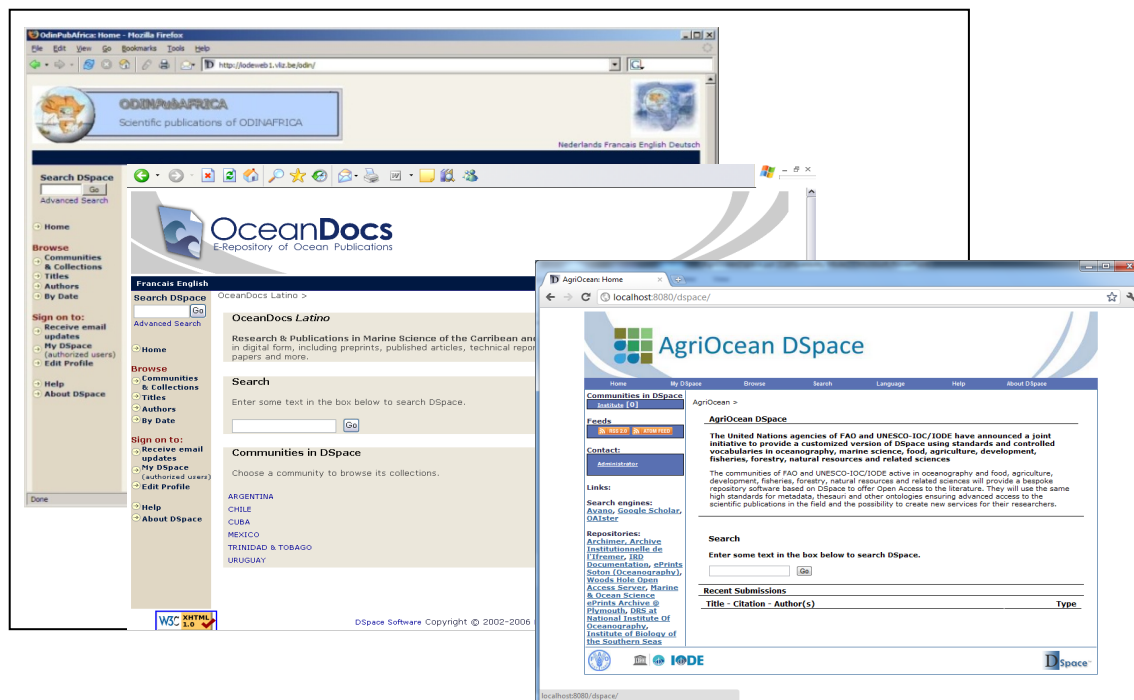


Fig. 1. From OdinPubAfrica to AgriOcean Dspace: Website in evolution (2004-2011).

The OceanDocs Network has now about 50 members. The OceanDocs³ repository includes more than 40 institutes from Africa and Latin America (3,654 docs). The new partners GEOHAB (Global Ecology and Oceanography of Harmful Algal Blooms) and ASCLME (Agulhas and Somali Current Large Marine Ecosystem) will start to submit their publications in the near future.

The Institute of Biology of the Southern Seas (IBSS – Sevastopol, Ukraine) started their repository in 2007 which is based on the OceanDocs implementation. Their participation in the software development was critical for the finalization of AgriOcean Dspace. The collection contains now 1830 documents⁴. IBSS is also responsible for the CEEMAR repository of OdinECET, containing 964 publications⁵.

ODINPimris and the National Institute of Oceanography(GOA, India) have created their repository using other software. ODINPimris uses Greenstone, while NIO uses a standard Dspace.

The OceanDocs network is set-up to support ODIN's (Ocean Data and Information Network) of IODE⁶ with the development of repositories and has the following goals:

- Facilitate publishing of research findings by scientists (e-journal as well as e-archive) thereby promoting the related research and increasing access by scientists to the international research forum, specifically for oceanography, marine and related sciences
- Make scientific publications in the field of marine science and oceanography more easily and freely accessible
- Enhance the internal scientific communication
- Use of common standards

³ <http://www.oceandocs.net>

⁴ <http://repository.ibss.org.ua/dspace>

⁵ <http://www.ceemar.org>

⁶ <http://www.iode.org>

Other projects are now getting linked to OceanDocs. Afrilib, the federated catalogue of OdinAfrica, will use OAI-compatible software, including AgriOcean Dspace, making it possible to exchange metadata between the local OdinAfrica partners, the Afrilib catalogue and the OceanDocs repository. The pilot project to archive data sets, Published Ocean Data, is the result of a cooperation between SCOR, MBLWHOI and IODE, and is evaluating the possibility to use AgriOcean Dspace too.

FAO and Dspace:

FAO is since 2008 interested in DSpace as a tool for AGRIS centres. They were looking for software to share information in a standardized manner, using more sophisticated and specialized metadata, a common semantics and interoperable syntaxes. The use of AGRIS AP as a metadata standard and the integration of AGROVOC as a controlled vocabulary were other requirements of FAO.

Based on version 1.5, Prof. dr. ARD Prasad of the Indian Statistical Institute (Bangalore, India) created Agris Dspace for FAO. He adapted the submission module to create richer metadata, adding for example the language option to a field. Agris AP was supported in the OAI-PMH module.⁷

The screenshot shows a web browser window titled 'Agris: Describe this Item - Mozilla Firefox'. The address bar shows 'http://193.190.8.15/agris/submit'. The form contains several input fields and a dropdown menu. The 'Identifiers' section has a dropdown set to 'ISSN' and an 'Add More' button. The 'Title' section has a text input field containing 'El Niño' and an 'Add More' button. The 'Supplementary Title' section has a text input field and an 'Add More' button. A language dropdown menu is open, showing options: Spanish, Thai, English, Spanish (highlighted), German, French, Italian, and Japanese. At the bottom, there is a field for 'Enter the names of the authors of this item below.' and a 'Klaar' button. The Zotero logo is visible in the bottom right corner.

Fig. 2. The submission module of Agris Dspace.

Kasetsart University, Bangkok (Thailand) has also developed a plug-in for the use of AGROVOC in Dspace 1.5.2. This beta-version has not been upgraded to the later versions of Dspace.

Cooperation between FAO and UNESCO-IOC/IODE

The communities supported by the United Nations agencies FAO and UNESCO-IOC/IODE are synergistic (Agris, ASFA and ODIN) and the standards on metadata (Agris AP) and controlled vocabularies (Agrovoc or ASFA thesaurus) are similar for both.

A common repository development is a logical result. Hasselt University Library has created for FAO and UNESCO-IOC/IODE a new version called AgriOcean DSpace which is available since 1 August 2011. It integrates the previous developments of both agencies in one customized version of DSpace.

2. AgriOcean DSpace: the best of both world

Release of AgriOcean DSpace 1.0

AgriOcean DSpace 1.0 was released mid August 2011 and is now available in source code at Google Code⁸ with the option to customize the lay-out and adapt the configuration file to local requirements. For partners with limited IT support, a Windows-based easy-to-install version is available.

The system is based on a type-based submission module and includes the possibility to enter more refined metadata, define the language on field level and use the DSpace authority control system for journals and the ASFA

⁷ http://www.dspace.org/images/Training_Materials/AGRIS%20AP%20add-on%20for%20DSpace.pdf

⁸ <http://code.google.com/p/agrioccean/>

and AGROVOC ontologies. AgriOcean DSpace focuses on Agris AP and MODS (OAI-PMH compliant), which are the most relevant formats for the communities involved.

AgriOcean DSpace

Home My DSpace Browse Search Language Help About DSpace

Communities in DSpace
 Africa [1537]
 GEOHAB [8]
 IODE [491]
 Latin America [1681]

Feeds
 RSS 2.0
 ATOM FEED

Contact:
 Administrator

Links:
 Search engines:
 Avano, Google Scholar,
 OAIster
 Repositories:
 Archimer, Archive
 Institutionnelle de
 l'Ifremer, IRD
 Documentation,
 ePrints Soton
 (Oceanography),
 Woods Hole Open
 Access Server, Marine
 & Ocean Science
 ePrints Archive @
 Plymouth, DRS at
 National Institute Of
 Oceanography,
 Institute of Biology of
 the Southern Seas

TEST DSpace instance >

AgriOcean DSpace

The United Nations agencies of FAO and UNESCO-IOC/IODE have announced a joint initiative to provide a customized version of DSpace using standards and controlled vocabularies in oceanography, marine science, food, agriculture, development, fisheries, forestry, natural resources and related sciences

The communities of FAO and UNESCO-IOC/IODE active in oceanography and food, agriculture, development, fisheries, forestry, natural resources and related sciences will provide a bespoke repository software based on DSpace to offer Open Access to the literature. They will use the same high standards for metadata, thesauri and other ontologies ensuring advanced access to the scientific publications in the field and the possibility to create new services for their researchers.

Search

Enter some text in the box below to search DSpace.

Recent Submissions

Title - Citation - Author(s)	Type
Beginning the Cuba-SDRP training and research collaboration. López, R.	Other
Nephrolithiasis and pyelonephritis in two west indian manatees (Trichechus manatus spp.) Journal of Wildlife Diseases, 44(3), p. 707-711 Keller, M. Moliner, J. L. Vásquez, G.	Journal Contribution
Situación de la pesquería del recurso almeja (Venus antiqua) en la X región. Investigación situación pesquerías bentónicas, 2006 Barahona T., N. González Y., J.	Report
Evaluación de desechos de pescado frescos y ensilados en la alimentación de híbridos de Clarias gariepinus x Clarias macrocephalus Revista cubana de investigaciones pesqueras, 27(1), p. 21-25 Llanes Iglesias, J. E. Toledo Pérez, J. Lazo de la Vega Valdez, J. M.	Journal Contribution
Desempeño reproductivo de Litopenaeus schmitti de cultivo en condiciones comerciales Revista cubana de investigaciones pesqueras, 27(1), p. 14-20 Pérez-Jar, L. Ramos Trujillo, L. Racotta Dimitrov, I. S.	Journal Contribution

FAO IODE DSpace

Fig. 3: The home page of AgriOcean Dspace.

The Enhanced Submission Module

In Dspace the interface of the submission module has only a limited range of options. These are defined by the 'initial questions': Published or not, more than one title and more than one attachment.

From experience, users need a more structured approach. Therefore, OceanDocs already included a type-based submission module. This module has been refined for AgriOcean Dspace and a new definition table has been created: input-forms-extended.xml.

```

<collection-type>
  <collection handle="default">
    <type name="Journal Contribution" />
    <type name="Book"/>
    <type name="Book Section" />
    <type name="Proceedings Paper" />
    ...
    <type name="Working Paper" />
    <type name="Map" />
    <type name="Data Set" />
    <type name="Other" />
  </collection>
</collection-type>

```

Fig. 4 Definition of type templates in input-forms-extended.xml

In the input-form-extended.xml file the different types are defined (see fig. 4) and for every type a different template can be created.

Fig. 5 Options in the submit module to select a template by type.

In AgriOcean Dspace, after choosing a collection, you can immediately start describing the document. If you need another template you can switch by selecting another type. A default type can be defined in the input-forms-extended file. Here it is 'Journal contribution'.

Fig. 6. Layout options in AgriOcean Dspace.

AgriOcean Dspace has included extra lay-out options compared to standard Dspace: grouping of fields (in a row and a block), defining field size and adding a language option for a field.

```

<fieldgroup>
  <fieldrow>
    ....
  </fieldrow>
  <fieldrow>
    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>bibliographicCitation</dc-
element>
      <dc-qualifier>volume</dc-qualifier>
      <label>Volume</label>
      <input-type size="4">onebox</input-type>
      <repeatable>>false</repeatable>
      <required></required>
    </field>
    ...
    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>bibliographicCitation</dc-
element>
      <dc-qualifier>endpage</dc-qualifier>
      <label>End page</label>
      <input-type size="4">onebox</input-type>
      <repeatable>>false</repeatable>
      <required></required>
    </field>
  </fieldrow>
</fieldgroup>

```

Fig. 7. Layout options defined in input-forms-extended.xml

The authority control of Dspace 1.7 has also been enhanced. The link between field and authority list is now defined in input-forms-extended (see Fig. 8). This way for every different type template a different authority list could be defined even on the same Dublin Core element.

```
<authority closed="true" editable="true" presentation="suggest"
limit="50">issn</authority>
```

Fig. 8. Use of authority control defined in input-forms-extended.xml.

For the journal titles, AgriOcean Dspace uses a table containing 37.000 titles. The ISSN is used as the authority code, which is also copied to the dc.identifier.issn field. Still it is possible to enter manually a missing title with its ISSN. Finally, the title, volume, issue, start and end page are kept in the database as separate values, but they are also concatenated to the dc.identifier.citation element. This has two advantages. AgriOcean Dspace keeps a very refined metadata, while the formatting of dc.identifier.citation is standardized.

Fig. 9 Authority control for Journal titles and ISSN.

The authority control system is used for the ASFA and AGROVOC keywords too. This authority control system gives only the opportunity to search through an alphabetic list of keywords using autosuggestion search options.

Fig. 10 Authority control for ASFA and AGROVOC.

The integration of a thesaurus and ontology system will be based on the work of Kasetsart University on Dspace 1.5.2. It will be integrated in the next version of AgriOcean Dspace.

Finally, the layout and authority control elements developed for the submission module are also used in the Edit interface of AgriOcean Dspace (See Fig. 11).

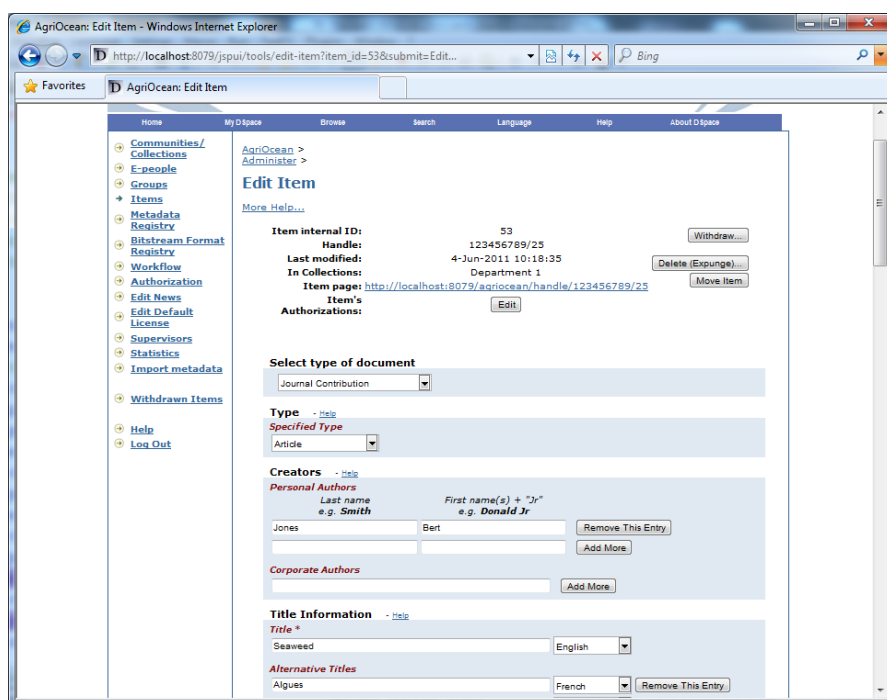


Fig. 11. Edit mode of AgriOcean Dspace.

Refinement of the OAI-PMH Output in AgriOcean Dspace

A main goal of AgriOcean Dspace is to create refined metadata and use richer standards like Agris AP and MODS.

The standard metadata format of Dspace is Qualified Dublin Core. Elements and qualifiers can be defined. It is also possible to create other formats, but because of the internal structure of Dspace, it will always be limited to a two level approach: element and qualifier. It has the advantage that it is easy to create new elements and qualifiers even out of the Dublin Core range. This is in not a problem since it is in fact only a way of structuring internally the metadata.

The main problem is how this internal structure is translated to the external world. In the first place how the metadata is exposed through the OAI-PMH protocol. The crosswalk system of Dspace is in that way very efficient, still it has its limitations (see further).

AgriOcean Dspace starts with very specific metadata fields. Extra elements were created when the the necessary granularity is missing: e.g. dc.bibliographicCitation.volume, dc.bibliographicCitation.issue, dc.bibliographicCitation.stpage, dc.bibliographicCitation.endpage. Basic elements like dc.identifier.citation are recreated through concatenation of these fields.

Through the system of crosswalks the output of Dublin Core and Qualified Dublin Core can easily be managed. AgriOcean Dspace enhances the crosswalk functionalities to support Agris AP and MODS fully.

Dspace collects values that are not defined in element and qualifier, like the language and the authority ID. These values could not be translated to a metadata format in the existing OAI-PMH module. Therefore the crosswalk system was rewritten.

First, It is now possible to define the concatenation of fields (which are not repeatable).

Example in MODS crosswalk:

```
dc.bibliographicCitation.conferencename = <name type="conference"><role><roleTerm
type="text">conference</roleTerm></role><namepart>%s,
$dc.bibliographicCitation.conferenceplace$s$,
$dc.bibliographicCitation.conferencedate$s$ </namepart></name>
```

Result in OAI-PMH, combining the elements conference name, place and date:

```
< name type="conference ">
  <role><roleTerm type="text">conference</roleTerm></role>
  < namepart >5th IAMSLIC Conference. Paris, 25 June 1999</ namepart >
</ name >
```

Figure 12. Concatenation of fields.

Secondly, the values language and authority can now also be translated.

Example in MODS crosswalk:

```
dc.subject.asfa = <subject authority="ASFAT" xml:lang="%l" URI="%a">%s</subject>
```

Result in OAI-PMH:

```
<subject authority="AGROVOC" xml:lang="en"
  URI="http://aims.fao.org/aos/agrovoc/c_2925">
  <topic>Fish meal</topic>
</subject>
```

Figure 13. Translation of values language and authority.

Thirdly, The xml tags are grouped. In Agris AP and MODS, deeper levels of XML are sometimes required. Therefore subtags are grouped under the tags of the higher level.

Example in AGRIS AP crosswalk:

```
dc.subject.asfa = <dc:subject><ags:subjectThesaurus scheme="ags:ASFAT"
  xml:lang="%l">%s</ags:subjectThesaurus></dc:subject>
dc.subject.agrovoc = <dc:subject><ags:subjectThesaurus scheme="ags:AGROVOC"
  xml:lang="%l">%s</ags:subjectThesaurus></dc:subject>
dc.subject.agrovoc-uri = <dc:subject><ags:subjectThesaurus
  scheme="ags:AGROVOC">%s</ags:subjectThesaurus></dc:subject>
```

Result in OAI-PMH:

```
<dc:subject>
  <ags:subjectThesaurus scheme="ags:ASFAT" xml:lang="en">Accumulation</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="ags:ASFAT" xml:lang="en">Fish meal</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="ags:ASFAT" xml:lang="en">Wastes</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="ags:ASFAT" xml:lang="en">Water masses</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="ags:AGROVOC" xml:lang="en">Fish meal</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="ags:AGROVOC" xml:lang="en">Wastes</ags:subjectThesaurus>
  <ags:subjectThesaurus
    scheme="ags:AGROVOC">http://aims.fao.org/aos/agrovoc/c_2925</ags:subjectThesaurus>
  <ags:subjectThesaurus
    scheme="ags:AGROVOC">http://aims.fao.org/aos/agrovoc/c_8307</ags:subjectThesaurus>
</dc:subject>
```

Figure 15. Grouping of xml tags.

But grouping can be turned off. At the end of the translation line the level of ungrouping is defined in Figure 16, below.

Example in MODS crosswalk:

```
dc.contributor.author= <name type="personal">
  <role><roleTerm type="text">author</roleTerm></role>
  <namePart>%s</namePart></name> name
```

Result in OAI-PMH:

```
<name type="personal">
  <role>
    <roleTerm type="text">author</roleTerm>
  </role>
  <namePart>Said, T.O.</namePart>
</name>
<name type="personal">
  <role>
    <roleTerm type="text">author</roleTerm>
  </role>
  <namePart>Hamed, M.A.</namePart>
</name>
```

Figure 16. Level of ungrouping.

The combination of granular internal metadata and a refined crosswalk system results in high metadata quality as can be seen in OceanDocs: For agris at http://www.oceandocs.org/odin-oai/request?verb=ListRecords&metadataPrefix=agris&set=hdl_1834_104, for mods at http://www.oceandocs.org/odin-oai/request?verb=ListRecords&metadataPrefix=mods&set=hdl_1834_104.

High quality metadata is not a goal in itself. It makes extra services possible if communities stick to the high quality metadata, which is proved by projects like VOA3R⁹.

⁹ <http://voa3r.eu/> - <http://voa3r.cc.uah.es>

Easy-to-install Version.

The Easy-to-install version was created to make AgriOcean Dspace available for smaller organizations with limited IT support. It is already distributed in OdinAfrica and will be tested in the Agris network.

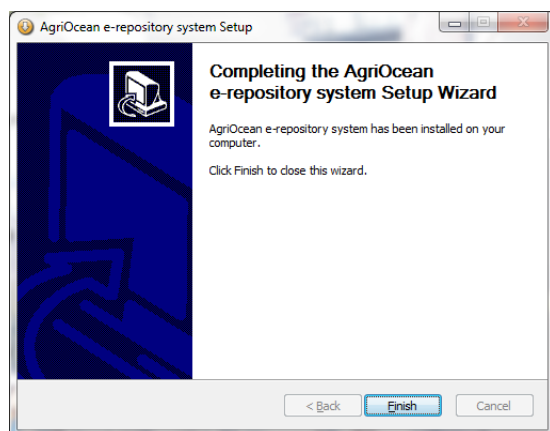
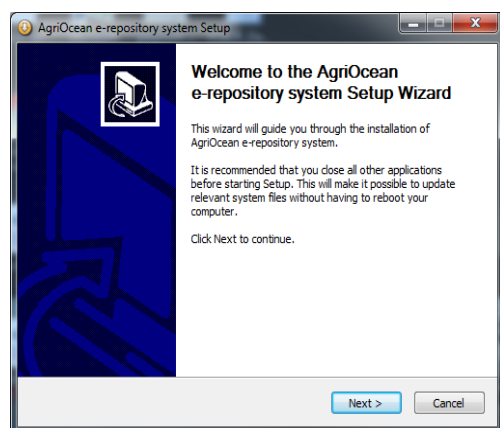


Fig. 17. The AgriOcean Easy Installer (10 min between both screenshots).

The package contains a builded AgriOcean Dspace and Postgresql. After installing Tomcat users simply have to define the main configuration elements of Dspace and Postgres during a set up procedure that takes about 15 minutes. The set-up procedure has been thoroughly tested for Windows Vista and 7, while a beta Windows XP version is also available¹⁰. The installer is created with the Open Source package 'SpringSource Tool Suite'.

3. Distribution and Support

AgriOcean Dspace is available at <http://code.google.com/p/agrioccean/downloads>. FAO created an AgriOcean DSpace community on AIMS, Agricultural Information Management Standards: <http://aims.fao.org/community/group/agrioccean-dspace>

AIMS¹¹ (Agricultural Information Management Standards) is a web portal managed by FAO of the United Nations that disseminates standards and good practices in information management to support the right to food, sustainable agriculture and rural development. AIMS supports the implementation of structured and linked information and knowledge by fostering a community of practice centered on the themes of interoperability, reusability and

¹⁰: http://code.google.com/p/agrioccean/downloads/detail?name=1_installation_agrioccean-dSpace-1_0-3.pdf

¹¹ <http://aims.fao.org/>

cooperation. It shares vocabularies, methodologies, tools and services in order to promote coherence in agricultural information.

On AIMS under ‘Tools’ a special section is dedicated to AgriOcean DSpace¹² containing background information, support material and the AgriOcean Dspace community pages. The community is set-up for all institutions that would like to install AgriOcean DSpace or are interested in the repository software. On its forum the AgriOcean DSpace team provides assistance with installing and using AgriOcean DSpace, answers questions and communicates new developments and solved bugs.

UNESCO-IOC/IODE will create complementary courses on the OceanTeacher platform¹³

¹² <http://aims.fao.org/tools/agrioccean-dspace> - <http://aims.fao.org/community/group/agrioccean-dspace>

¹³ <http://classroom.oceanteacher.org/>

4. Future Developments of AgriOcean Dspace.

The batch import functionality will be available at the beginning of December 2011. It will then be possible to upload metadata files in Agris AP, Endnote and Zotero-RIS format.

Authority control will be extended to author names and the thesaurus/ontology plug-in will make it easier to manage ASFA, Agrovoc and even geographical descriptors. This will be included in the next release which is planned for August 2012. Next releases will keep in pace with the Dspace version available at that moment.

AgriOcean Dspace is part of the set of tools to be distributed among the VOA3R data providers for the implementation of document repositories. The possibilities of Linked Open Data (LOD) for AgriOcean Dspace will be explored.

OceanDocs and AgriOcean Dspace.

AgriOcean Dspace is already installed on the OceanDocs repository and partners like IBSS (Ukraine), CEEMAR (Eastern Europe) and KMFRI (Kenya).

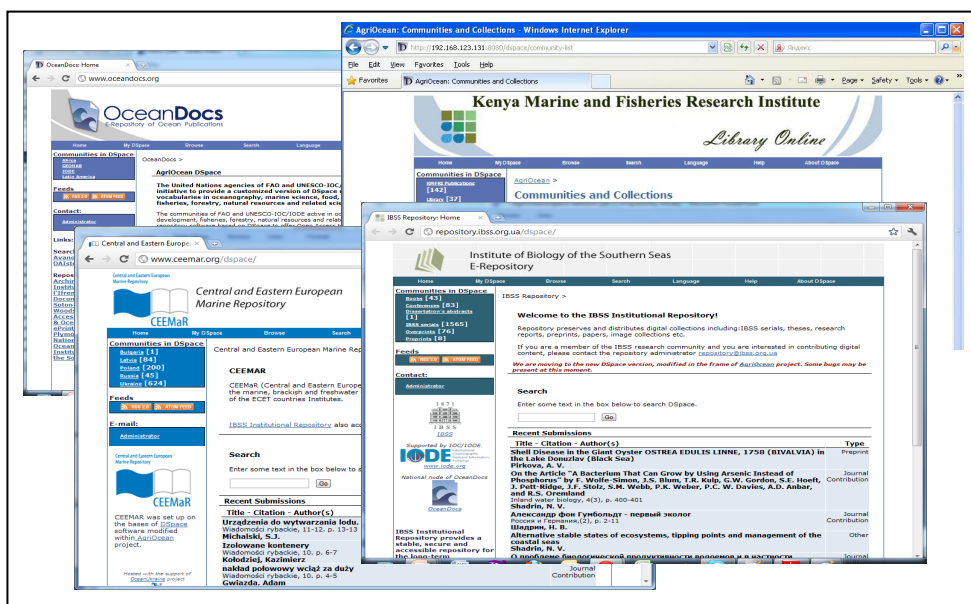


Fig. 13. Installations of AgriOcean Dspace.

IODE partners will have more options in developing their repository collection. They can still join the OceanDocs repository, but they can use now the software package available at the AgriOcean Google Code site. Organizations with limited IT capacity will opt for the Easy-to-install version, while others can use the source code to have a more customized repository.

The OceanDocs repository will harvest the IODE-related repositories. In the future, it will also be possible to harvest metadata and full text, creating a mirror of the local repositories, which not always have a stable environment. OceanDocs is already harvested by AVANO and will be by Agris of FAO.